J Mol Evol (1986) 23:313–319

Journal of
Molecular Evolution

Divergence and Evolution of Geminivirus Genomes

Alan J. Howarth[1] and Robert M. Goodman[2]

[1] Department of Plant Pathology, University of Arizona, Tucson, Arizona 85721, USA
[2] Calgene, Inc., 1920 Fifth Street, Davis, California 95616, USA

Summary. The nucleic acid sequences of three geminiviruses with bipartite genomes and of two viruses having a single genome component were analyzed and phylogenetic relationships deduced. Sequences in coding and noncoding regions were considered at the nucleotide and amino acid levels by several methods. The results suggested that the viruses are phylogenetically related to different degrees. All the viruses contain in an intergenic region a consensus sequence (TAATATTAC) that is postulated to be required for a critical virus function, such as replication and/or transcription. Estimates of divergence in one putative gene that all of the viruses share were used to construct a phylogenetic tree. Among the bipartite-genome viruses, bean golden mosaic virus and tomato golden mosaic virus are more closely related than either is to cassava latent virus. The single-component viruses (maize streak and wheat dwarf viruses) and one of the two DNA components of the other three viruses were postulated to be distant relatives descended from a common ancestral sequence.

Key words: Geminivirus — Viral evolution — Bean golden mosaic virus — Tomato golden mosaic virus — Cassava latent virus — Maize streak virus — Wheat dwarf virus

Introduction

Geminiviruses are plant viruses having morphologically unique geminate or twinned particles with genomes of circular single-stranded DNA (Matthews 1982). Geminiviruses have long been known to be important agriculturally (Goodman 1981) and have recently attracted interest as potential vectors for plant genetic engineering (Howarth and Goodman 1982; Buck and Coutts 1983). This interest is especially keen in the search for vectors in monocotyledonous plants, which some geminiviruses naturally infect. Complete nucleotide sequences are now available for four viruses and two strains of a fifth (Stanley and Gay 1983; Hamilton et al. 1984; Howell 1984; Howarth et al. 1985; MacDowell et al. 1985; Mullineaux et al. 1984).

Bean golden mosaic virus (BGMV; Howarth et al. 1985), cassava latent virus (CLV; Stanley and Gay 1983), and tomato golden mosaic virus (TGMV; Hamilton et al. 1984) are serologically interrelated (Sequeira and Harrison 1982; Stein et al. 1983; Roberts et al. 1984) and are transmitted in nature by a whitefly (*Bemisia tabaci*). These viruses each have a genome composed of two different DNA molecules; each molecule is about 2600 nucleotides long and both are essential for disease development (Hamilton et al. 1983; Morinaga et al. 1983; Stanley 1983). Six genes have been postulated (Howarth et al. 1985), but only one, that for the capsid protein, has been experimentally verified (Townsend et al. 1985).

Each of the geminiviruses with two-component genomes has an approximately 200-nucleotide sequence that is identical (in BGMV) or almost identical (in CLV and TGMV) on each DNA molecule of the virus. This sequence, called the common region, is very different from virus to virus, except for a similar inverted repeat sequence located near the 3′ end of each common region. These inverted re-

Offprint requests to: A.J. Howarth

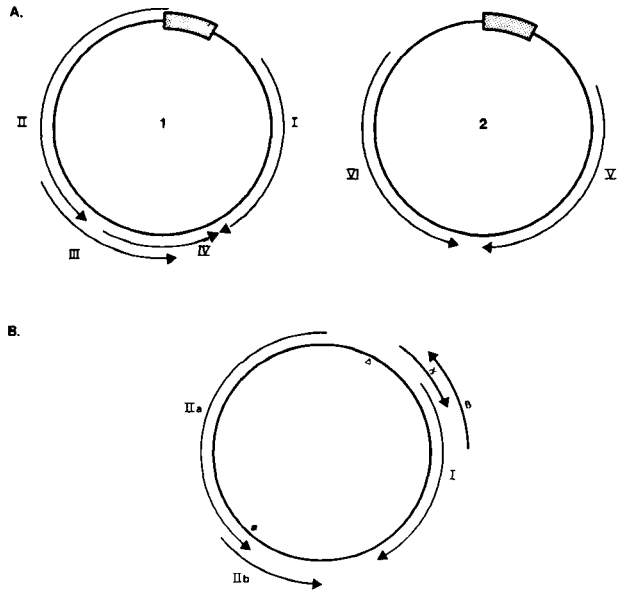

Fig. 1A, B. Maps of geminiviral open reading frames (ORFs). Nucleotide sequences of geminiviruses with two DNA components have the organization shown in **A.** Arrows represent conserved ORFs, which are designated by Roman numerals. Stippled boxes represent the 200-nucleotide common regions. Geminiviruses with one DNA component are organized as in **B.** The designation of ORFs I, IIa, and IIb signifies the relationship inferred with ORFs I and II in DNA 1 of the bipartite viruses. The hairpin that contains the sequence TAATATTAC is marked △. The putative amber codon proposed by Howell (1984) is marked ◆.

peats may form stable hairpin structures with 11- to 12-base-pair (bp) GC-rich stems and 11- to 12-nucleotide AT-rich, single-stranded loops. The conservation of common-region sequences in a given virus suggests an important function for the entire 200-nucleotide region (Stanley and Gay 1983; Hamilton et al. 1984; Howarth et al. 1985). The sequences of the common regions of the different viruses are so different that it suggests that some virus-specific property, such as viral host range, may reside in that difference.

Wheat dwarf virus (WDV; MacDowell et al. 1985) and two isolates of maize streak virus (MSV; Howell 1984; Mullineaux et al. 1984) have been characterized and they apparently each contain only one DNA molecule, of about 2600–2700 nucleotides.

Relationships among viruses and their evolution have historically been largely speculative subjects (Matthews 1985). Certain obvious conclusions could be deduced about the viral families Reoviridae and Rhabdoviridae because they include viruses that infect vertebrates, invertebrates, and plants (Matthews 1982). However, the increasing availability of nucleotide and amino acid sequences and methods for analyzing them now invite comparisons across viral taxa. For example, interesting sequence homologies have been detected between a plant comovirus and animal picornaviruses (Argos et al. 1984; Franssen et al. 1984); among plant caulimoviruses and animal retro- and hepadnaviruses (Toh et al. 1983); and among a plant bromovirus, tobamovirus, alfalfa mosaic virus, and animal alphaviruses (Haseloff et al. 1984; Ahlquist et al. 1985).

We have compared the sequences in the conserved open reading frames and in the noncoding regions of several geminiviruses; from analyses of these sequences we derive estimates of divergence of putative genes, and suggest some evolutionary implications.

Methods

We used four methods of scoring the similarity of nucleotide and amino acid sequences that had been aligned with computer assistance or visually. Unitary matrix scoring assigned a score of 1 or 0 to identical or different residues being compared, respectively. These scores were summed and divided by the maximum score to obtain the percentage of similarity. Amino acid sequences compared in this way were considered also for their conservative amino acid sequence homologies. Amino acid substitutions were given a score of 1 when substituted amino acids were from the same group. Groups of amino acids were C; F, W, and Y; Q, E, N, and D; H, K, and R; L, M, and V; and A, G, P, S, and T. We used the method of Perler et al. (1980) to calculate the percentage divergences (corrected for multiple mutations) in silent and replacement sites within codons. Silent and replacement sites are sites at which nucleotide changes result in synonymous codons or amino acid replacement substitutions, respectively. The method of McLachlan (1971) scores the structural similarities of likelihood of interchange among amino acids. The method of Fitch and Margoliash (1967) scores the mutation distance, which is the minimum number of nucleotide changes in a codon required to convert an amino acid into the amino acid with which it is aligned. We used mutation-distance data to construct phylogenetic trees (Fitch and Margoliash 1967).

Results

Geminiviruses with Bipartite Genomes

Comparison of open reading frames (ORFs) among BGMV, CLV, and TGMV sequences reveals that six ORFs are present in all three viruses (Hamilton et al. 1984; Howarth et al. 1985). These ORFs show varying degrees of sequence similarity, but are organized identically, in the three viruses (Fig. 1). Conserved ORFs occur in both viral and complementary strands.

We compared the conserved ORF sequences among the three viruses and calculated their divergence (Table 1). Also, we used the putative amino acid sequences of the proteins that may be encoded by these ORFs to compute their protein similarities (Table 2) and their minimum mutation distances (data not shown). The several methods yielded similar results, as a comparison of methods for analysis of the similarity of BGMV and TGMV sequences shows (Table 3).

Table 1. Percentage corrected divergences between nucleotide sequences[a]

Comparison	Replacement sites						Silent sites					
	I	II	III	IV	V	VI	I	II	III	IV	V	VI
BGMV–TGMV	6.9	19.9	21.8	15.8	21.4	8.2	142.2	113.2	48.2	55.4	95.4	157.8
BGMV–CLV	26.6	29.8	44.2	49.6	86.9	44.3	199.7	175.4	80.8	103.0	130.4	172.1
TGMV–CLV	26.4	29.6	46.4	47.9	84.0	46.4	171.0[b]	144.0	89.2	103.0	137.2	165.0[b]

BGMV, bean golden mosaic virus; CLV, cassava latent virus; TGMV, tomato golden mosaic virus

[a] I–VI refer to open reading frames (ORFs) as given in Fig. 1. Divergences were calculated according to Perler et al. (1980)

[b] Some operations in the calculation of silent-site divergence were mathematically undefined. In those cases the terms in question were not corrected for multiple mutations and the divergence values are necessarily underestimates. See Perler et al. (1980)

Table 2. Amino acid sequence similarities among bipartite geminiviruses[a]

Comparison	Open reading frame					
	I	II	III	IV	V	VI
BGMV–TGMV	93.4	83.9	82.8	87.7	83.4	93.7
BGMV–CLV	83.9	79.4	70.9	69.3	56.5	67.6
TGMV–CLV	84.7	78.6	70.6	69.0	57.1	71.5

[a] I–VI refer to ORFs as given in Fig. 1. Similarities were calculated according to McLachlan (1971). Abbreviations as in Table 1

Table 3. Comparison of methods for analysis of BGMV and TGMV sequences[a]

ORF	Method			
	Amino acid sequence homology (unitary matrix)	Conservative amino acid sequence homology[b]	Method of McLachlan (1971)	Method of Perler et al. (1980)[c]
I	90	95	93.4	93.1
II	71	85	83.9	80.1
III	68	83	82.8	78.2
IV	76	86	87.7	84.2
V	70	85	83.4	78.6
VI	90	94	93.7	91.8

Abbreviations as in Table 1

[a] Numbers indicate percentage similarity

[b] Percentage of total amino acid residues encoded by each ORF that are either identical or conservation substitutions

[c] Numbers obtained by subtracting the percentage divergence in replacement sites from 100

At the broadest level, the genes in DNA 1 are more highly conserved than those in DNA 2. BGMV and TGMV are more closely related to each other than either is to CLV. Also, BGMV and TGMV are about equally divergent from CLV.

As expected, different proteins accumulate mutations at different rates. Gene I, which on the basis of recent evidence is thought to encode the capsid protein (Townsend et al. 1985), is the most highly conserved gene; gene V is the most highly diverged. Genes II, III, and IV are interesting cases because they have regions of mutually overlapping sequences in different reading frames (Stanley and Gay 1983; Hamilton et al. 1984; Howarth et al. 1985). Gene II is probably little affected, because its overall length (1059 nucleotides) is large compared with its overlap with gene III (250 nucleotides). The sequence of gene III, however, overlaps significantly with that of gene IV. Genes III and IV have relatively high levels of divergence in their replacement sites and relatively low levels in their silent sites. Low levels of divergence in silent sites may be the result of the existence of fewer neutral sites (Perler et al. 1980): Nucleotide changes at neutral sites in one gene may be at replacement sites in an overlapping gene, and hence selected against. Surprisingly, high levels of divergence were calculated in replacement sites. This observation appears counterintuitive because selection presumably was acting on two proteins whose genes happen to occupy much of the same sequence, rather than on only one protein. One would expect, therefore, very low levels of divergence.

Gene VI appears to be composed of two domains, one of which is relatively well conserved and the other which is relatively diverged. The conserved domain is approximately the 5′-terminal two-thirds of the BGMV and CLV genes and the entire TGMV gene. TGMV gene VI is truncated by a termination codon and would yield a protein of about 21.1 kilodaltons, versus 33.1 and 33.6 kilodaltons for BGMV and CLV, respectively (Stanley and Gay 1983; Hamilton et al. 1984; Howarth et al. 1985). The divergent domain is the 3′-terminal one-third of the BGMV and CLV genes.

Geminiviruses with Monopartite Genomes

The DNA sequences of the Kenyan (MSV-K; Howell 1984) and Nigerian (MSV-N; Mullineaux et al. 1985) strains of MSV are about 98% homologous, differing by only 59 changes out of about 2680 nucleotides. The papers presenting these sequences displayed their viral-strand sequences in opposite orientations. The orientation of MSV-N is probably the correct one, because it is consistent with the

	Stem	Loop	Stem
BGMV 1 and 2	GGCGGCCATCCG	ATATAATATTAC	CGGATGGCCGCC
CLV 1 and 2	GGGGCCAACCG	TATAATATTAC	CGGTTGGCCCC
TGMV A	GCGGCCATCCG	TTTAATATTAC	CGGATGGCCGC
TGMV B	GCGGCCATCCG	TTTTAATATTAC	CGGATGGCCGC
MSV-N	GCAGGAAAAGAAGGCGCG	ACTAATATTAC	CGCGCCTTCTTTTCCTGC
MSV-K	GCAAGAAAAGAAGGCGCG	CACTAATATTAC	CGCGCCTTCTTTTCCTGC
WDV	GGGGGCCTCCACGCGGG	TTATAATATTAC	CCCGCGTGGTGGCCCCC
Consensus		TAATATTAC	

Fig. 2. Sequences of conserved hairpins. Sequences of inverted repeats are displayed to show which sequences may base pair in stems of hairpins and which would be in the single-stranded loops. The sequence TAATATTAC is in the loop of each hairpin and is adjacent to stem sequences on its 3′ side, whereas 2–5 residues separate the conserved loop sequence from the stem sequences on its 5′ side. BGMV, bean golden mosaic virus; CLV, cassava latent virus; MSV-K, maize streak virus, Kenyan strain; MSV-N, maize streak virus, Nigerian strain; TGMV, tobacco golden mosaic virus; WDV, wheat dwarf virus

orientation of the sequence of WDV DNA (MacDowell et al. 1985) and of the whitefly-transmitted geminivirus genomes (see below). Both reports of MSV DNA sequences mention heterogeneity in their sequences. For the purposes of our comparison, we considered only the predominant sequences and not the possible substitutions. All references to nucleotide position are according to the numbering for MSV-N.

Only 5 of the 59 differences in sequence are due to insertions or deletions; the remainder are merely nucleotide substitutions. All but one of the insertions or deletions occur in putative intergenic regions. The one nucleotide insertion/deletion that is in a possible coding region of MSV-K is an extra G residue between nucleotides 1504 and 1505. The effect of that insertion is to place MSV-Ks ORFs P1a and P1b (which correspond to the ORFs designated IIa and IIb in Fig. 1) in the same triplet reading frame. Howell (1984) suggested that readthrough of the amber codon terminating ORF P1a is required for expression of ORF P1b. Because ORFs IIa and IIb (Fig. 1) of MSV-N and WDV, however, are in different triplet frames, we tend to favor the inference that these genes are expressed independently and that the MSV-N sequence is correct.

We examined other positions within codons at which differences had been reported between MSV-K and MSV-N (data not shown). Because most of the differences between the strains occur in the third positions of codons, we concluded that ORFs I, IIa, and IIb are probably real genes (see also Mullineaux et al. 1985). An ORF in MSV-N that may encode a protein of 21.8 kilodaltons is probably not a real gene because its putative counterparts in MSV-K and WDV are interrupted by termination codons: A nucleotide change in the codon formed by residues 716–714 results in a TGA termination. We cannot tell from this comparison if ORFs 10.9, 11.2, and 13.0 in MSV-N are real genes or not because there is not enough difference between the two sequences for third-position changes or termination codons to indicate which ORFs, if any, are merely fortuitous. Comparison of MSV to another leafhopper-transmitted geminivirus, WDV (MacDowell et al. 1985), showed that all of the ORFs in WDV had counterparts in MSV-N; however, no counterpart for ORF 11.2 in MSV-N was detected in WDV. Convincing alignments were presented for the putative products of ORFs I and IIa + IIb. We quickly aligned WDV ORF 10.1 with MSV-N ORF 10.9, assisted by a homology matrix program (Pustell and Kafatos 1984). We named this apparently conserved ORF "alpha." Using the same program we also found weak similarity between WDV ORF 14.5 and MSV-N ORF 13.0. These ORFs, which we call "beta," overlap in the complementary strand positions occupied by ORFs 1 and alpha in the viral strand (Fig. 1).

Sequence Analysis in the Common or Large Intergenic Regions

The sequences in the stem and loop portions of the inverted repeats found in the common regions of the bipartite viruses differ only slightly among the three viruses (Fig. 2). The sequence TAATATTAC is conserved in the loops of all three viruses. It is interesting that this sequence resembles the TATA box sequence of eukaryotic transcriptional pro-

moters (McKnight and Kingsbury 1982). The functional significance of this resemblance is unknown; however, the conservation of inverted repeat sequences suggests control of essential viral functions such as initiation of DNA replication, promotion or termination of transcription, or protein binding.

Both of the MSV sequences and the WDV sequence contain inverted repeats that potentially form hairpin or stem–loop structures. Hairpin 1 of MSV-N (Mullineaux et al. 1984) and hairpin 2 of WDV (MacDowell et al. 1985) attracted our attention because they are in intergenic regions and because of the similarity of the sequences in their single-stranded loops to the loop sequences found in the common regions of the bipartite geminiviruses (Fig. 2).

These hairpins lie in large intergenic regions between the putative coat protein gene (gene I), transcribed in the clockwise direction, and gene II, transcribed in the counterclockwise direction (Fig. 1). These results suggest that MSV and WDV DNAs are organized similarly to DNA 1 of the bipartite geminiviruses.

The similarity between these two groups is further reinforced by the observation that the consensus sequence TAATATTAC is found in loops of hairpins in all six sequenced geminiviruses. The stem sequences of hairpin 1 of MSV and hairpin 2 of WDV are quite dissimilar and are very different from the sequences found in the stems of BGMV, CLV, and TGMV. Thus, it is probably the sequence TAATATTAC in the loop of the stable hairpin and possibly the topological position of the stem–loop structure that is important in the function of this distinctive feature of the geminiviruses. This sequence seems to be important regardless of the number of DNAs in the genome, the insect vector, or the host specificity of the virus.

Discussion

The geminate morphology exhibited by virions of geminiviruses is unique in nature, as far as we know. It is very curious indeed to discover that within this unique collection of plant viruses are two distinct groups. One group is naturally transmitted by leafhoppers, apparently has genomes composed of 2.6 kb of single-stranded DNA, and infects mono- or dicotyledonous plants. The other group of geminiviruses is transmitted by *B. tabaci,* has genomes of two 2.6-kb single-stranded DNAs, and infects dicots.

These two groups of geminiviruses have, besides their virion morphology, the following characteristics in common. First, all geminivirus genomes that have been sequenced have the sequence TAATATTAC located on a single-stranded loop of a

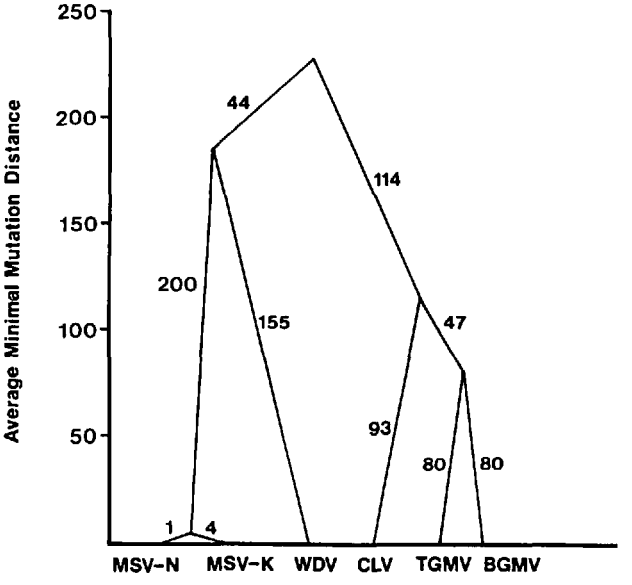

Fig. 3. Phylogenetic relationships among geminiviruses. The numbers along the branches of the tree represent the corrected mutation distances between genes II of the viruses, which were aligned as in Fig. 8 of MacDowell et al. (1985), with the BGMV sequence data similarly aligned. Each apex is placed at an ordinate value representing the average of the sums of all mutations in the lines of descent from that apex. The percentage standard deviation of this tree is 0.54. Abbreviations as in Fig. 2

stable hairpin that is in an intergenic region of the genome. Second, in DNA 1 of the bipartite viruses and in the DNA of the monopartite viruses, this hairpin is situated between the putative coat protein gene (gene I) and a long ORF (gene II) that is conserved in the two geminivirus subgroups.

Kikuno et al. (1984) detected homology between genes I (encoded on DNA 1) and V (encoded on DNA 2) of CLV, and we have observed the same relationship in BGMV and TGMV (data not shown; see also Rao, 1985). This result suggested to them that the bipartite genome evolved from a monopartite ancestor. We agree with that assessment and, furthermore, conclude that the geminiviruses with mono- and bipartite genomes are ancestrally related. We have used our estimates of divergence in gene II, the most highly conserved geminiviral gene, to construct an evolutionary tree (Fig. 3). This tree is based on mutation distance (Fitch and Margoliash 1967). Perler et al. (1980) found that mutation rates in replacement sites of codons were linear over the past 500 million years in preproinsulin and globin genes, though the results on the evolution of superoxide dismutase genes (Lee et al. 1985) suggest that one must use caution when interpreting divergence data. However, assuming linear rates of change over time, the monopartite–bipartite divergence is estimated to have taken place twice as long ago as the divergence of BGMV or TGMV from their common ancestor with CLV, and three times as long ago as the divergence between BGMV and TGMV.

The relative closeness of the relationships among the bipartite viruses may reflect the several features they have in common. In addition to the postulated divergence to a bipartite genome, these viruses have the same insect vector and are found in plants grown in similar subtropical environments. The greater phylogenetic distance between the monopartite viruses may be consistent with differences between the viruses. Their vectors belong to different genera. MSV infects maize and other grasses (but not wheat) in tropical and subtropical Africa, while WDV is found in the temperate climate of northern Europe.

Among the RNA viruses of plants, multipartite genomes are common. Several groups of animal RNA viruses also have segmented genomes. The mechanisms by which segmented viral genomes arise in nature are unknown. Similarities in genomic organization between polioviruses (nonsegmented) and comoviruses (bipartite) suggest that the two-component comovirus genome may have arisen from a single-component precursor (Matthews 1985). A similar evolutionary route to segmentation can be inferred from analysis of homologies in nonstructural proteins (Haseloff et al. 1984; Ahlquist et al. 1985; Matthews 1985) shared among tobacco mosaic virus (nonsegmented) and brome mosaic and alfalfa mosaic viruses (which both have three genomic RNAs). Our conclusions about geminivirus phylogeny suggest another route for the origin of a segmented genome. In this case, divergence from a common progenitor sequence appears to have occurred, rather than segmentation of a larger progenitor sequence.

Acknowledgments. We thank Lydia Delis for assistance with illustrations and Christine Facciotti for preparation of the manuscript. A.J.H. acknowledges financial support from the Arizona Agricultural Experiment Station, project ARZT-174758-H-05-12.

References

Ahlquist P, Strauss EG, Rice CM, Strauss JH, Haseloff J, Zimmern D (1985) Sindbis virus proteins nsP1 and nsP2 contain homology to nonstructural proteins from several RNA plant viruses. J Virol 53:536–542

Argos P, Kamer G, Nicklin MJH, Wimmer E (1984) Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. Nucleic Acids Res 12:7251–7267

Buck KW, Coutts RHA (1983) The potential of geminiviruses in the construction of plant host vector systems. Plant Mol Biol 2:351–354

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Franssen H, Leunissen J, Goldbach R, Lomonossoff G, Zimmern D (1984) Homologous sequences in non-structural proteins from cowpea mosaic virus and picornaviruses. EMBO J 3: 855–861

Goodman RM (1981) Geminiviruses. In: Kurstak E (ed) Handbook of plant virus infections and comparative diagnosis. Elsevier/North Holland, Amsterdam, pp 879–910

Hamilton WDO, Bisaro DM, Coutts RHA, Buck KW (1983) Demonstration of the bipartite nature of the genome of a single-stranded DNA plant virus by infection with the cloned DNA components. Nucleic Acids Res 11:7387–7396

Hamilton WDO, Stein VE, Coutts RHA, Buck KW (1984) Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: potential coding regions and regulatory sequences. EMBO J 3:2197–2205

Haseloff J, Goelet P, Zimmern D, Ahlquist P, Dasgupta R, Kaesberg P (1984) Striking similarities in amino acid sequence among nonstructural proteins encoded by RNA viruses that have dissimilar genomic organization. Proc Natl Acad Sci USA 81:4358–4362

Howarth AJ, Goodman RM (1982) Plant viruses with genomes of single-stranded DNA. Trends Biochem Sci 7:180–182

Howarth AJ, Caton J, Bossert M, Goodman RM (1985) Nucleotide sequence of bean golden mosaic virus and a model for gene regulation in geminiviruses. Proc Natl Acad Sci USA 82:3572–3576

Howell SH (1984) Physical structure and genetic organisation of the genome of maize streak virus (Kenyan isolate). Nucleic Acids Res 12:7359–7375

Kikuno R, Toh H, Hayashida H, Miyata T (1984) Sequence similarity between putative gene products of geminiviral DNAs. Nature 308:562

Lee YM, Friedman DJ, Ayala FJ (1985) Superoxide dismutase: an evolutionary puzzle. Proc Natl Acad Sci USA 82:824–828

MacDowell SW, Macdonald H, Hamilton WDO, Coutts RHA, Buck KW (1985) The nucleotide sequence of cloned wheat dwarf virus DNA. EMBO J 4:2173–2180

Matthews REF (1982) Fourth report of the International Committee on Taxonomy of Viruses. Intervirology 17:1–199

Matthews REF (1985) Viral taxonomy for the nonvirologist. Annu Rev Microbiol 39:451–474

McKnight SL, Kingsbury R (1982) Transcriptional control signals of a eukaryotic protein-coding gene. Science 217:316–324

McLachlan AD (1971) Tests for comparing related amino acid sequences: cytochrome c and cytochrome c 551. J Mol Biol 61:409–424

Morinaga T, Ikegami M, Miura K (1983) Infectivity of the cloned DNAs from multiple genome components of bean golden mosaic virus. Proc Jpn Acad 59:363–366

Mullineaux PM, Donson J, Morris-Krsinich BAM, Boulton MI, Davies JW (1984) The nucleotide sequence of maize streak virus DNA. EMBO J 3:3063–3068

Mullineaux PM, Donson J, Stanley J, Boulton MI, Morris-Krsinich BAM, Markham PG, Davies JW (1985) Computer analysis identifies sequence homologies between potential gene products of maize streak virus and those of cassava latent virus and tomato golden mosaic virus. Plant Mol Biol 5:125–131

Perler R, Efstratiadis A, Lomedico P, Gilbert W, Kolodner R, Dodgson J (1980) The evolution of genes: the chicken preproinsulin gene. Cell 20:555–566

Pustell J, Kafatos FC (1984) A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. Nucleic Acids Res 12:643–655

Rao JKM (1985) Sequence homology among the coat proteins of geminiviruses. Biochem Biophys Res Commun 130:892–896

Roberts IM, Robinson JD, Harrison BD (1984) Serological relationships and genome homologies among geminiviruses. J Gen Virol 65:1723–1730

Sequeira JC, Harrison BD (1982) Serological studies on cassava latent virus. Ann Appl Biol 101:33–42

Stanley J (1983) Infectivity of the cloned geminivirus genome requires sequences from both DNAs. Nature 305:643–645

Stanley J, Gay MR (1983) Nucleotide sequence of cassava latent virus DNA. Nature 301:260–262

Stein VE, Coutts RHA, Buck KW (1983) Serological studies on tomato golden mosaic virus, a geminivirus. J Gen Virol 64:2493–2498

Toh H, Hayashida H, Miyata T (1983) Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature 305:827–829

Townsend RJ, Stanley J, Curson SJ, Short MM (1985) Major polyadenylated transcripts of cassava latent virus and location of the gene encoding coat protein. EMBO J 4:33–37

Received February 8, 1986/Revised and accepted March 20, 1986

Note added in proof: The nucleotide sequence of beet curly top virus DNA has recently appeared (Stanley J, Markham PG, Callis RJ, Pinner MS (1986). The nucleotide sequence of an infectious clone of the geminivirus beet curly top virus. EMBO J 5:1761–1767). Cloned DNA from this leafhopper-transmitted geminivirus is infectious and consists of a single molecular of 2993 nucleotides. Three of its genes closely resemble genes II, III, and IV of the whitefly-transmitted geminiviruses in both sequence and organization. However, the putative coat protein gene is more similar to coat protein genes of other leafhopper-transmitted geminiviruses (MSV and WDV).